



ELSEVIER

Journal of Food Composition and Analysis 17 (2004) 557–564

JOURNAL OF
FOOD COMPOSITION
AND ANALYSIS

www.elsevier.com/locate/jfca

Report

Food and nutrient changes: software designed to enhance data quality

Ellen Anderson*, Lois C. Steinfeldt, Jaspreet K.C. Ahuja

Food Surveys Research Group, Beltsville Human Nutrition Research Center, Agricultural Research Service, USDA, BARC-West, Bldg. 005, Beltsville, MD 20705, USA

Received 9 July 2003; received in revised form 20 February 2004; accepted 3 March 2004

Abstract

Food and nutrient databases must be continually updated to reflect changes in the food supply and improvements in data values. The Food Databases Management System (FDMS) is a software application designed to manage the US Department of Agriculture's Food and Nutrient Database for Dietary Studies used for food consumption surveys. Staff of the Food Surveys Research Group use FDMS to accomplish a variety of complex database management tasks. FDMS was created using ObjectPal, the programming language of the commercially available database program Paradox[®]. With FDMS, users can update the database files with new information such as adding a new food code. Another important feature is the ability to integrate time-related changes into the database files. Better data values, due to improved analytical procedures and sampling methods, can replace older values in the database. When there is a true change to the food product itself, FDMS allows users to specify these time-specific changes with start dates and end dates. Programmed into FDMS is a series of automated quality control checks which monitor processes as they occur. Interfile quality control checks are also in place so that related files can be synchronized and updated. FDMS also contains historical information which staff can use when setting priorities for database updates.

Published by Elsevier Ltd.

Keywords: Database management; Food and nutrient trends; Software development; Quality control

1. Introduction

The United States Department of Agriculture (USDA) has been monitoring the food usage and nutrient intakes of Americans for decades in support of its mission to ensure a safe and nutritious

*Corresponding author. Tel.: +1-301-504-0174; fax: +1-301-504-0377.

E-mail address: eanderson@rbhnrc.usda.gov (E. Anderson).

national food supply. To properly monitor food and nutrient consumption patterns in a population, the dietary database used for analyses should correctly reflect the food supply for a given period of time. The variety and dynamic nature of national food supplies make the maintenance of the database a challenge. New food products, changes in package sizes, and food formulation modifications in established brands must be monitored, as well as improvements in analytical techniques which yield better data values. Once important changes in the food supply and improved data values are identified, these changes must be incorporated into the database in an accurate and timely manner. The Food Surveys Research Group of the USDA recognizes the need for efficient methods of updating and maintaining its complex food and nutrient database used for food consumption surveys and other dietary research. The Food Databases Management System (FDMS) was developed in response to this requirement, as a tool to help ensure a high-quality database.

2. Overview of USDA's Food and Nutrient Database for Dietary Studies (FNDDS)

The Food Surveys Research Group of USDA maintains a large, complex database that consists of three main components: food descriptions, food portions and weights, and nutrients ([US Department of Agriculture, 2000](#)). The database is designed to track changes and facilitate analysis of intake trends in the United States. Changes made to the database are categorized as either data improvements or food changes ([Anderson et al., 2001](#)). Data improvements are changes in data values generated from improved analytical techniques or better food sampling methods. Improved analytical data for several key foods in the US food supply generated by the National Food and Nutrient Analysis Program (NFNAP) ([Haytowitz et al., 2002](#)) were recently incorporated into the database. One example is potassium in hamburger/hot dog rolls, which was found to be overestimated by approximately 30%. New values from data improvements overwrite old values; in other words, they are applied retroactively in the database. A food change occurs when a food itself actually changes, either in nutritive content or in package size. The old value remains in the system with dates attached indicating the time frame for which the value is valid. Likewise, the new value has a date that specifies when it goes into effect. One example of an important food change is the enrichment of grain products with folate as required by the Food and Drug Administration in 1998.

There are over 7700 food descriptions in the main food descriptions file, including foods discontinued from the market which are tracked for historical purposes. Each of these food items has a set of gram weight measures and nutrient values attached to it. Each food code has a recipe that contains ingredient items from the USDA Nutrient Database for Standard Reference ([US Department of Agriculture, 2002](#)). An in-house nutrient file for these recipe ingredients tracks food changes and data improvements. For example, the ingredient code for the cereal Kellogg's All Bran[®] has three values for iron which are valid for three different periods of time. The USDA's Nutrient Data Laboratory (NDL) provides these nutrient values. Food descriptions are used to identify the food item. Keeping food descriptions and brand names updated in the database makes searching and selecting the correct food code easier during the coding of survey data. The same advantages come from updating food weights and measures in the database. However, unlike changes to food descriptions, changes to food

weights may impact nutrient intake. Thus, it is important to be able to track the changes in sizes of manufactured food packages and also to apply improved gram weight measurements to the database.

The complexity and size of the database necessitate careful quality control practices. Along with the main 7700 food description records, the database contains approximately 30,000 g weight records, 4300 portion descriptions, and 7200 additional food descriptions that describe brand names or other details pertinent to the main food descriptions. Each food code has values for 61 nutrients and other components, totaling more than 420,000 data records in the survey nutrient database. Many of the database files are interrelated by common fields. Thus, coordinated effort is required to keep related records among the various data files synchronized as changes are made throughout the database.

3. Description of the Food Databases Management System (FDMS): main features

FDMS was designed with five basic goals in mind: (1) to enter and edit changes to the food descriptions, food portion and weights, and nutrient files in an accurate and efficient manner; (2) to organize complex management tasks and quality control checks affecting the database into one on-line system; (3) to provide users easy access to the database and related statistical and historical data for information gathering and decision making; (4) to be accessible and easy to use for individuals with little or no database management or programming experience; (5) to be modular in format so that system enhancements can be continuously and seamlessly added over time. FDMS was created using ObjectPal, an object-based, event-driven, visual programming language that is part of the commercially available database software package Paradox[®] (Corel Corporation and Corel Corporation Limited, 1999). The system essentially is a collection of on-line forms containing objects such as buttons, text boxes, and drop down lists that provide information or allow the user to accomplish a task related to the food databases. The FDMS opening screen displays a menu of the main modules of the system (Fig. 1). Brief descriptions of these modules are listed in Table 1.

3.1. Access levels

Each button on the main menu is considered a separate module that directs the user to distinct functions and database information. Three levels of access exist to control user rights to modules. The basic user level of access does not require a password and permits activities such as data entry to suggest database changes and browsing historical information. The intermediate level is reserved for data reviewers. In this access level, the data reviewer checks all database changes proposed by the basic users for accuracy and correctness. The data reviewer has the authority to change or delete entries made by other users and gives final approval for all changes affecting the database. The highest level of access is for the database manager. Tasks such as running the final quality control checks and the incorporation of approved changes into the database are done by the database manager. Access to tasks performed by the data reviewer and database manager is controlled by passwords. No special expertise in database management software or programming is necessary for any user access

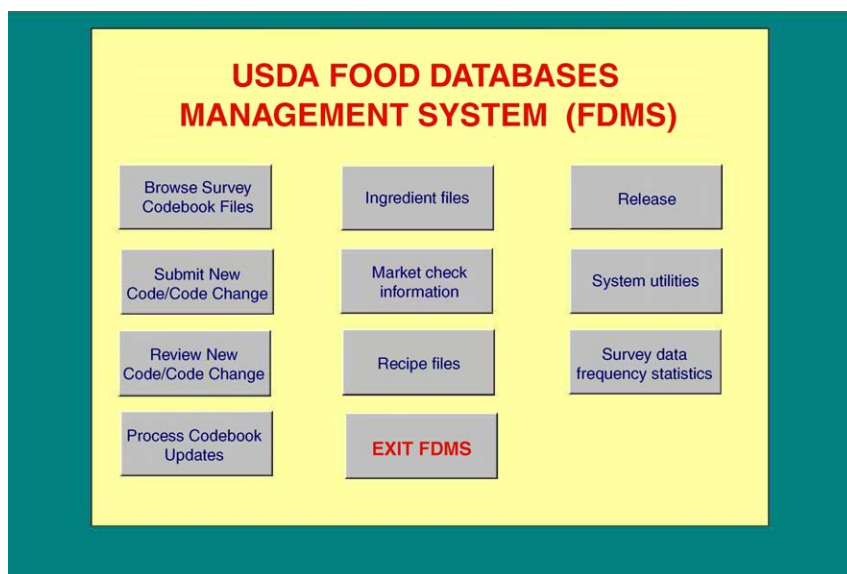


Fig. 1. FDMS opening screen.

Table 1

Summary of main modules of the Food Databases Management System (FDMS)

Module name	Functions
Browse survey codebook files	Read-only search of current food descriptions and weights, proposed database changes, and historical files; basic user access
Submit new code/code change	Enter/edit proposed database changes; basic user access
Review new code/code change	Enter/edit/approve proposed database changes; intermediate level access
Process codebook updates	Final quality control checks and incorporation of changes into master database files; highest level access
Ingredient files	Read-only search of current and past, proposed changes to, and historical files of nutrient files; enter/edit/approve proposed database changes; quality control checks for input files from NDL and incorporation into master files; mixed level access
Market check information	Search product information, enter/edit market check request; mixed level access
Recipe files	Generate files for recipe manager, run interfile checks; intermediate level access
Release	Export database files for public release and other applications; highest level access
System utilities	Backup functions, copy files; highest level access
Survey data frequency statistics	Read-only survey frequency data for food codes, modifications, and gram weight records; basic level access

level in FDMS. The different access levels exist primarily to manage the flow of data more efficiently. The following section provides more details on how the access levels are used in FDMS.

3.2. *Entering revisions to the database*

Users navigate through FDMS modules to locate the appropriate data entry form for the type of change being suggested. Before any revisions are entered, the user must locate the specific food code that will be affected by the change. Depending on the type of change and the database component involved, the data entry form displays the required fields in which users enter the appropriate information for a proposed change. For example, if the user wants to suggest a change to the cup weight of potato chips, FDMS will display a form that prompts the user to specify the type of change (data improvement or food change), the new gram weight, and a comment documenting the reason for the change. Many fields, such as the food code number, food description, and date of entry, are pre-filled by FDMS so the user does not need to enter them manually. All data entry forms contain a comment field in which users may enter text to serve as documentation for the proposed change. Users may maneuver between fields by using the keyboard Enter key or the mouse. FDMS utilizes a set of transaction data files to store users' suggested changes to the database. Using transaction data files in this manner preserves the content of the master database files, also called multi-version database, until the proposed changes are designated correct and valid.

For the food descriptions component and the food portions and weights component (also known in-house as “the codebook”), each suggested change that is entered in FDMS is assigned a progress status which indicates its stage of processing. Basic users assign a progress status of “Pending” if the proposed change is still being researched and the user is not ready to release it to the data reviewer. A progress status of “Submit” indicates that the proposed change is ready to be reviewed by the data reviewer. After the data reviewer checks the suggested change, the progress status may be marked as “Approved” if the change is deemed valid and all data entries are correct and ready for the database update. If the data reviewer determines the proposed change is not ready for incorporation into the multi-version files, then the progress status would be marked as “Hold”. A progress status of “N/A” means the data reviewer believes the proposed change is not necessary and should not become part of the multi-version database. Proposed changes with a status of “Hold” or “N/A” are saved in the system for reference in case similar changes are suggested in the future. Using predefined status labels on proposed changes has proven to be a useful feature in the management of database updates.

Revisions to the ingredient files are usually entered into FDMS in batch mode using data files supplied by NDL. All data files undergo a series of checks to verify that the file structures are correct and that the changes are valid and complete. The nutrient data files are also checked against nutrient parameters such as the sum of proximate components, difference between the sum of fatty acid classes and total fat, and the difference between vitamin A (RAE) and retinol. Reports of values exceeding pre-set limits are printed and reviewed. The database manager has the option of adding all the data in the batch file or selecting data to be added into transaction files. The data are held in these files until the change is deemed valid and correct.

For the food descriptions component, the following types of changes may be made using FDMS: add a new food code or discontinue an existing food code; change a food description associated with a food code; add, discontinue or change an additional food description associated with the main food description. Revisions that affect the food portions and weights component

include: add, discontinue, or change a gram weight associated with a food code; create a new portion description or discontinue an existing portion description; change a portion description. The following types of changes to the ingredient files may be entered with FDMS: add a new ingredient; discontinue an existing ingredient; change a description; change nutrient values; add new nutrients for existing ingredient codes; add, discontinue or change gram weights associated with an ingredient code.

3.3. *Quality control checks*

FDMS contains several types of automatic checks and data entry aids on the on-line forms which help to ensure high-quality information. Examples of some of these features include verification that a new food code number or a new ingredient number is unique, automatic assignment of sequential line numbers for new additional food descriptions and gram weight records, verification that start dates and end dates assigned to a food change are valid and synchronized, and checking that required fields in a record are not blank. Certain changes made to one file may have an effect on another file. FDMS performs interfile checks automatically and alerts the user of the situation via a message box shown on the screen. For instance, if a food code or ingredient code is marked to be discontinued, FDMS queries the recipe file to determine if the proposed discontinued code is used as an ingredient in a current recipe. Likewise, a portion description cannot be discontinued if it is currently used in a gram weight record. Having automatic checks built into FDMS helps to keep the files in the relational database synchronized with each other and improves communication among database management staff.

Before the master files for the food descriptions and food portions and weights components are updated with new and changed records, the database manager runs a final series of quality control checks from the “Process Codebook Update” module on the FDMS main menu. These checks are performed only on records with a progress status of “Approved” set by the data reviewer. If a problem is found in the Approved records, a detailed report listing the food code in question and the type of error is printed. The database manager must correct any errors before continuing with the database update. FDMS creates backup files of the “Approved” records and the original multi-version database files as a safeguard before the final update begins. Similar procedures are also followed for updating nutrient values for ingredient files.

In the process of running the database update, FDMS generates reports for the number and type of changes being applied to the master database files which can be used as an additional operational check of the system. New codes and new components to food codes and ingredient codes generate new records in the master database files. Discontinued items result in a change to the end dates of the records. New data values from data improvements overwrite the old values in the particular records, while new values from food changes are added as new records with specific effective dates applied to both the new and old values. In addition, FDMS stores historical data in a special database each time a change is applied to the multi-version database. The historical database is useful to staff who want to research when, why, and how many changes were implemented in the multi-version database over a period of time. FDMS coordinates all of the database file changes during this final step of the database update in a matter of minutes and is transparent to the user.

3.4. Browsing the databases and data frequency tables for information gathering

Other important features of FDMS are the browse and survey data frequency statistics modules. The browse modules give users read-only access to certain database files. Users may search the current multi-version codebook files by food code, food description, additional food description, and portion description (Fig. 2). If users want to know what database changes have been proposed and are awaiting data review, they may search the transaction database by the type of change. Often, users need to know the background of a previous database enhancement. Browsing the historical files gives users access to database updates that have occurred since 1993.

Like the food descriptions and the food portions and weights files, the ingredient files can be accessed in a similar fashion within FDMS. Users may search on the current multi-version files or single-version files, or a series of pre-1994 data files. Proposed changes to the in-house nutrient file such as new nutrient codes and changes in nutrient values, and past changes documented in the historical files may be browsed as well.

Before the database can be updated, USDA nutritionists must research the changes and improvements affecting the database and set priorities as to the types of changes being considered. One way to determine where to focus efforts for a database update is to review survey frequency data. This FDMS module provides information on the number of times a food code or modification was reported in past surveys. Users may sort the food codes based on most

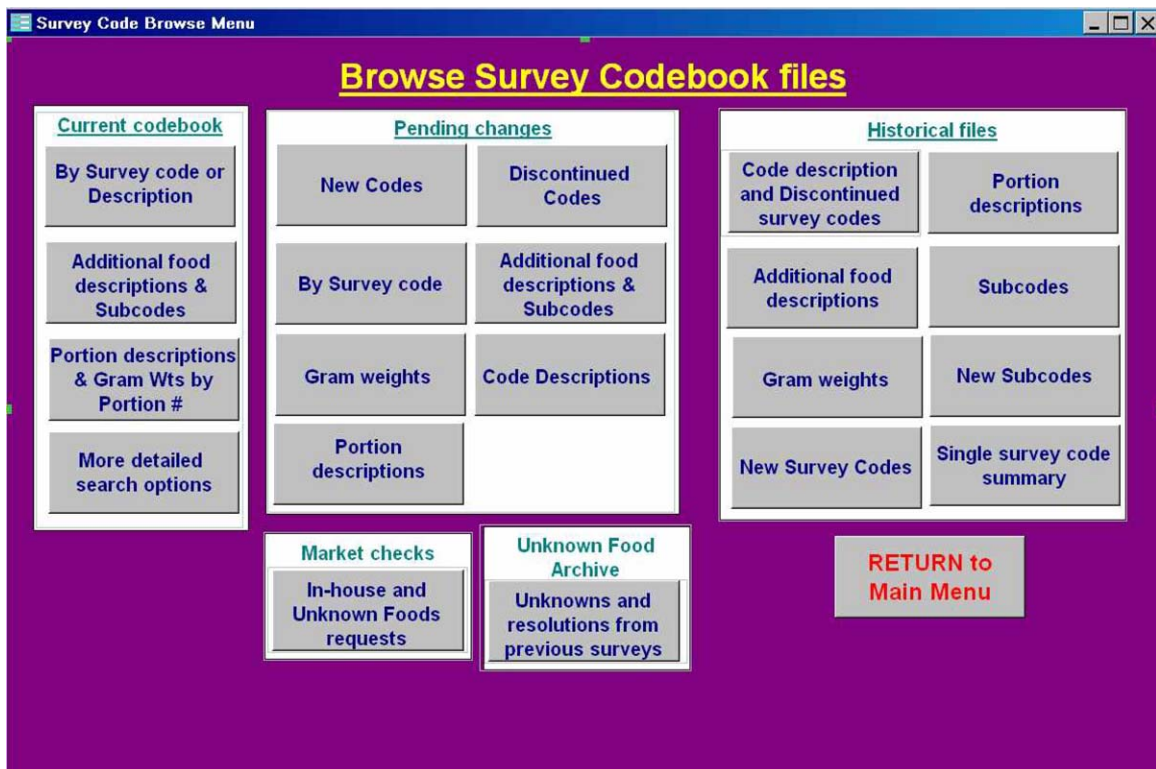


Fig. 2. Browse menu for food descriptions and food portions and weights files.

frequently reported for a given survey, or the frequency may be broken down by year. Also, the survey frequency of individual gram weights associated with a food code is included. The information provided in the data frequency module has proven to be a valuable and much-used resource by staff.

4. Conclusions

Processing time and efficiency for updating the Food and Nutrient Databases for Dietary Studies have greatly improved since FDMS has been in use. With FDMS, proposed changes are entered on-line via a user interface that has data entry aids and built-in error checking. Quality control checks are programmed at all stages of processing, allowing correction of errors as soon as possible. The time required to run final quality control checks, add the changes to the master database files, and create the historical records for the update is about 30 min for a typical update using FDMS.

Although the FDMS has been operational at the Food Surveys Research Group since May 2001, it is still a “work-in-progress” that was designed to expand over time to meet new challenges. Some of the major future enhancements planned for FDMS include linking an image database of food label, nutrient, and gram weight documents (e.g., market check information) for reference through FDMS to aid in documenting database updates and reduce the reliance on hard copy files; adding historical information about hard-to-code foods reported in our surveys; and more querying and report printing capability in the browse modules.

5. Disclaimer

Mention of commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture over others not mentioned.

References

- Anderson, E., Perloff, B., Ahuja, J.K.C., Raper, N., 2001. Tracking nutrient changes for trends analysis in the United States. *Journal of Food Composition and Analysis* 14, 287–294.
- Corel Corporation and Corel Corporation Limited, 1999. ObjectPal for Paradox[®], version 9. Ottawa, Ontario.
- Haytowitz, D.B., Pehrsson, P.R., Holden, J.M., 2002. The identification of key foods for food composition research. *Journal of Food Composition and Analysis* 15, 183–194.
- US Department of Agriculture, 2000. Agricultural Research Service. Continuing Survey of Food Intakes by Individuals 1994–96, 1998. CD-ROM, National Technical Information Service, Accession No. PB2000-500027.
- US Department of Agriculture, 2002. Agricultural Research Service. USDA Nutrient Database for Standard Reference, Release 15. Nutrient Data Laboratory Home Page, <http://www.nal.usda.gov/fnic/foodcomp>